

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<https://hdl.handle.net/2066/224631>

Please be advised that this information was generated on 2021-11-02 and may be subject to change.



Challenges in Investigating the Effective Components of Feedback from Routine Outcome Monitoring (ROM) in Youth Mental Health Care

Maartje A. M. S. van Sonsbeek^{1,2} · Giel J. M. Hutschemaekers^{1,3,4} · Jan W. Veerman⁴ · Ad Vermulst⁵ · Marloes Kleinjan^{6,7} · Bea G. Tiemens^{1,3,4}

Accepted: 31 August 2020 / Published online: 24 September 2020
© The Author(s) 2020

Abstract

Background Studies on feedback in youth mental health care are scarce and implementation of feedback into clinical practice is problematic.

Objective To investigate potentially effective components of feedback from Routine Outcome Monitoring (ROM) in youth mental health care in the Netherlands through a three-arm, parallel-group, randomized controlled trial in which a literature-based, multi-faceted implementation strategy was used.

Method Participants were randomly allocated to three conditions (basic feedback about symptoms and quality of life; basic feedback supplemented with clinical support tools; discussion of the feedback of the second condition with a colleague while following a standardized format for case consultation) using a block randomization procedure, stratified by location and participants' age. The youth sample consisted of 225 participants (mean age = 15.08 years; 61.8% female) and the parent sample of 234 mothers and 54 fathers (mean age of children = 12.50 years; 47.2% female). Primary outcome was symptom severity. Secondary outcomes were quality of life and end-of-treatment variables. Additionally, we evaluated whether being Not On Track (NOT) moderated the association between condition and changes in symptom severity.

Results No significant differences between conditions and no moderating effect of being NOT were found. This outcome can probably be attributed to limited power and implementation difficulties, such as infrequent ROM, unknown levels of viewing and sharing of feedback, and clinicians' poor adherence to feedback conditions.

Conclusions The study contributes to our limited knowledge about feedback from ROM and underscores the complexity of research on and implementation of ROM within youth mental health care.

Trial registration Dutch Trial Register NTR4234 .

Keywords Routine outcome monitoring · Feedback · Youth mental health care · Randomized controlled trial · Implementation

✉ Maartje A. M. S. van Sonsbeek
m.van.sonsbeek@propersona.nl

Extended author information available on the last page of the article

Background

Worldwide 10–20% of children and adolescents experience mental disorders (World Health Organization 2018). Mental disorders are the leading cause of disability in young people, and they severely limit young people's development, educational attainments, and opportunities to lead fulfilling and productive lives as adults. Therefore, providing successful treatments for mental disorders and prevention of treatment failures are the main goals of clinical practice. However, many children, adolescents, and adults drop out of therapy prematurely (Swift and Greenberg 2012; Warnick et al. 2012) or have negative therapy outcomes (Reese et al. 2014; Warren et al. 2010). Nevertheless, clinicians are overly optimistic about their effectiveness (Walfish et al. 2012), and they are unable to predict which clients are likely to deteriorate (Hannan et al. 2005; Hatfield et al. 2010). Clients who do not achieve positive change during treatment will require a disproportionately greater amount of treatment resources (Lambert et al. 2007). Clinicians, therefore, may benefit from having systematic and reliable information about the functioning of their clients (Hamilton and Bickman 2008).

Routine Outcome Monitoring (ROM) refers to the feedback of information from regular assessments of clients' progress throughout treatment, to clinicians or clients, or both, to facilitate clinical decision making (Scott and Lewis 2015). The feedback indicates what components of treatment do and do not seem to be working so that the clinician can, through dialogue with the client, be more responsive to deterioration and modify the treatment when needed in order to reduce treatment failures (Bickman et al. 2012). The increasing body of evidence showing that feedback can improve outcomes prompted the American Psychological Association Taskforce on Evidence Based Treatments (2006) to recommend that clinicians routinely collect and use client-report data to inform treatment (Sparks and Duncan 2018).

The effects of feedback in adult mental health care have been studied extensively, and results have indicated small negative to large positive effects (Carlier et al. 2012; Gondek et al. 2016; Knaup et al. 2009; Lambert et al. 2018; Østergård et al. 2018). Positive effects have been found on both intervention processes (e.g., improved client–clinician relationships and clients' engagement in treatment) and outcomes (e.g., reduction of symptoms, shorter treatment length, and fewer deteriorated clients). Feedback appeared to be especially effective when it included clinical support tools (CST; Harmon, et al. 2005), which are practical suggestions for overcoming possible obstacles (e.g., problems with self-efficacy, the social network, motivation, and the therapeutic alliance; Krägeloh et al. 2015; Shimokawa et al. 2010) in order to achieve a good outcome. The best results have been reported for clients who had not been improving during treatment, the so-called Not On Track (NOT) clients (De Jong et al. 2014; Simon et al. 2012). However, the most recent Cochrane review (Kendrick et al. 2016) and meta-analysis (Lambert et al. 2018) found very small differences in outcome between the feedback and no-feedback groups. The additional analyses in these papers suggested that feedback might improve outcomes for NOT clients and reduce the number of sessions required for OT clients. Again, the differences were small ($d=0.22$ and $d=0.33$, respectively). Moreover, the evidence for all comparisons was graded as low quality, because all included studies were considered at high risk of bias. In most cases this was due to inadequate blinding of assessors, significant attrition at follow-up, and indirectness of the evidence.

Despite the fact that providing continuous feedback has been viewed as part of good practice and feedback studies in adult mental health care are abundant, the implementation

of ROM is often problematic and the take-up by clinicians is low (Patterson et al. 2006). Surveys spanning different countries indicate that fewer than 20% of practitioners (17.9% of psychiatrists, 11.1% of psychologists, and 13.9% of masters-level practitioners) engage in ROM, and as few as 5% use it during every session (Lewis et al. 2019). Jensen-Doss et al. (2018) showed that only 13.9% of clinicians reported using standardized progress measures at least monthly and 61.5% never used them. Even in RCTs, as many as half of the clinicians do not use feedback during treatment (Simon et al. 2012). Good implementation into clinical practice seems crucial for feedback to have an effect on outcome (Brattland et al. 2018). It seems necessary that implementation efforts also target clinicians' attitudes, motivation and skills through training and supervision and that clinicians have the opportunity to discuss feedback with their clients and allow it to inform the treatment.

Research concerning feedback in youth mental health care is scarce. A recent systematic review and meta-analysis (Tam and Ronan 2017) identified only 12 studies that could be included. The results suggested that collection and application of continuous feedback from youth clients throughout treatment, particularly on a session-by-session basis, could produce positive, short-term effects on youth well-being. However, the effects of the feedback were small. There has been only one Cochrane review (Bergman et al. 2018), and it reported only six published randomized controlled trials, which were mostly with older children and adolescents (11–18 years old). Five of these studies were conducted in the United States and one in Israel. The studies showed only small negative to small positive effects on clients' symptoms, treatment acceptability (a reduction in drop-out), the therapeutic alliance, duration of treatment, number of treatment sessions, and client satisfaction. Additionally, Bickman et al (2011), (2016) found that better implementation of the feedback intervention resulted in better outcomes and that effects were stronger when clinicians viewed more feedback reports (a dose–response effect). Nevertheless, no firm conclusions could be reached about the effectiveness of feedback in psychological therapy for children and adolescents, because of considerable inconsistency in results among different studies and because the evidence was considered of very low-certainty because of a high risk of bias (Bergman et al. 2018). None of the studies was blinded or attempted to blind clinicians and participants (high risk of performance bias); only one study had blind outcome assessors (high risk of detection bias); and all of the studies had incomplete outcome data and/or non-transparent reporting of participants' flow through the study (high risk of attrition bias). It was concluded that additional studies are needed that avoid the risk of performance, detection and attrition biases, also include younger children and are conducted in countries other than the United States.

Besides the paucity of studies on feedback in youth mental health care, little is known about mechanisms that underlie the effects of feedback and the specific components of the feedback. Most theories about the mechanisms of feedback start by comparing current treatment results with the final goal of recovery (cf. Contextualized Feedback Intervention Theory Riemer and Bickman 2011). This results in a positive or negative evaluation of the clinician's performance. It is expected that clinicians who receive a negative evaluation will change their performance if they prefer externally generated feedback (de Jong et al. 2012); and/or they accept the feedback and have sufficient self-efficacy to adjust the treatment they deliver accordingly (cf. Goal-Setting Theory; Locke and Latham 1990); and/or they have detailed action plans regarding when, where, and how their behavior should change (Ivers et al. 2010). Research on the components of feedback has suggested that feedback should be specific, tailored to the needs and preferences of the clinician, written or graphic instead of verbally delivered, delivered directly after data collection to concretize the connection between the feedback and the

clinician's behavior, given frequently so that changes in processes and outcomes can be clarified directly and necessary corrective actions can be taken, and supplemented with concrete suggestions and directive interventions regarding ways to improve (CST; Harmon et al. 2005; Miller et al. 2005; Seidman et al. 2010; Slade et al. 2008). Above all, feedback can have an effect only when clinicians actually pay attention to and actually use the feedback (Claiborn and Goodyear, 2005).

In brief, research on feedback in youth mental health care is scarce, little is known about the mechanisms underlying the effects of feedback, and good implementation of the feedback into clinical practice has been problematic. The purpose of this study was to extend the literature on the effective components of feedback in youth mental health care while optimizing the chance that clinicians would actually implement the feedback in their day-to-day practice. We investigated several potentially effective components of feedback from ROM, in youth mental health care, and in the Netherlands. We extracted the feedback conditions directly from feedback theories, and we also examined a mechanism through which feedback might work. Additionally, we used a literature-based, multi-faceted strategy to implement the feedback conditions in clinical practice. We compared three different feedback conditions. In the first condition, clinicians received basic feedback regarding clients' symptoms and quality of life. In the second condition, the feedback from the first condition was extended with clinical support tools. Thus, the feedback in the second condition was more specific and more tailored to the needs of the clinicians, and it provided more information about the clinicians' actual behavior. In the third condition, the feedback from the second condition had to be discussed with a colleague while following a standardized format for case consultation. We thereby prevented clinicians from disregarding the feedback, and a peer who had personal relevance for the clinician provided additional advice. In all conditions, children, adolescents, and their parents were asked to complete the questionnaires. The primary outcome measure was symptom severity, and the secondary outcome measures were quality of life and end of treatment variables. Overall, we expected that the effects of the feedback would increase with the intensity with which it was provided. Hence, it was hypothesized that the third feedback condition would be most effective in decreasing symptom severity and improving clients' quality of life, and that it would affect the end of treatment variables most favorably. We also examined the role of being Not On Track (NOT) and hypothesized that feedback would be most effective for children and adolescents who were NOT.

Method

Design

We examined the effects of different components of feedback from ROM in youth mental health care through a three-arm, parallel-group, randomized controlled trial, in which a literature-based, multi-faceted implementation strategy was used (van Sonsbeek et al. 2014). We also examined the role of being NOT. The results were analyzed separately for children and adolescents (youth sample) and their parents (parent sample). The study was registered in the Dutch Trial Register (NTR4234) and was approved by the Ethics Committee of Radboud University's Faculty of Social Sciences (ECG2012-1304-031). The authors declare that they have no competing interests.

Procedure

The study was conducted at all four outpatient youth departments of a specialized mental health care institution in the eastern part of the Netherlands. The study was depicted as a way to better implement ROM and to investigate different types of feedback in order to determine which feedback worked best for improving clients' outcome.

All families with children and adolescents between the ages of 4 and 17 years who were referred to one of the youth departments from January 2014 to December 2015 were approached about participating in the study. The families were approached through a flyer that was included with the registration forms. No restrictions were made regarding the kind of problem (e.g., developmental, anxiety, or mood disorder) or the kind of treatment (e.g., individual vs. group treatment, cognitive-behavioral vs. solution-focused treatment, frequent vs. sporadic treatment, treatment for the child or adolescent only vs. one with additional training in parenting skills). The only exclusion criterion was having an insufficient understanding of the Dutch language. For children between 4 and 11 years old, only parents were approached about participating. For adolescents between 12 and 17 years old, both the adolescent and his or her parents were invited to participate. Adolescents younger than 16 years were approached about participating only after parental consent had been obtained. In accordance with the medical ethical guidelines, all parents and 16 and 17 year-old adolescents provided written informed consent within the registration forms. Adolescents younger than 16 years provided written assent within the registration forms. The secretaries at the youth departments checked the registration forms for completeness. Subsequently, the children, adolescents, and their parents were added to the electronic health record system at the mental health institution.

Inclusion to the study was performed twice per week. The primary researcher (MvS) and the research assistant received overviews with all referred families, double checked whether the registration forms had been completed and whether the children, adolescents, and parents had agreed to be included in the study. Children, adolescents, and parents who agreed to participate were added to the research file and were randomly assigned to one of the three conditions. The participants who were assigned to the second or the third feedback condition were asked to complete additional web-based questionnaires. Participants were followed from January 2014 to August 2016.

All of the participants were asked to complete web-based questionnaires at the start of treatment (baseline assessment), one and a half months after treatment had started, subsequently every three months during treatment, and at the end of treatment. Participants in the first condition filled out the standard set of questionnaires (i.e., Strengths and Difficulties Questionnaire, KIDSCREEN, and Youth Thermometer), and participants in the second and third feedback conditions filled out one additional questionnaire (Treatment Support Measure).

During the course of the study, approximately 170 different clinicians worked at the youth departments. Due to institutional changes, the total number of clinicians and the number of clinicians assigned to each department varied during the course of the study. In January 2014, 76 clinicians worked in the youth departments. Data on clinicians' age, sex, and discipline were missing for 15 (19.7%), 0 (0.0%), and 2 (2.6%) of the cases, respectively. The mean age of the remaining clinicians was 47.41 years ($sd=10.98$; range, 27–63) and 78.9% ($n=60$) of them were female. The clinicians' primary disciplines were psychology ($n=36$; 48.6%), nursing ($n=20$; 27.0%), social work ($n=5$;

6.8%), psychiatry ($n=5$; 6.8%), non-verbal therapies (e.g., psychomotor therapy; $n=5$; 6.8%), and education ($n=2$; 2.7%), and one of them was a student ($n=1$; 1.4%). The clinicians could access the feedback about the results from the questionnaires directly in the ROM system or one day after they had been completed in the patient's electronic health record. Before the study began, all clinicians were trained to use feedback from ROM. The primary researcher organized one or two training sessions in each department, which were scheduled on days and times when most clinicians were already present. Each manager communicated that the training for his or her team was obligatory, and he or she was present at the training. Accordingly, with a few exceptions, each clinician was trained. During the training, the clinicians received information about ROM, the feedback from ROM, and the study, and they practiced with interpreting the feedback, discussing the feedback, and using the case consultation form.

The clinicians also received an implementation package with information about ROM, the study, and interpretation of the feedback. The implementation package was discussed during the training, and it could be used as both a reference when the feedback was discussed and as a reminder to check for new feedback and to discuss the feedback with a colleague.

Throughout the study, the primary researcher organized multiple update meetings in each department, both at the request of the managers and on her own initiative. These meetings, like the training sessions, were scheduled at times when most of the clinicians were already present and were required by their managers to attend. The managers themselves were also present at the update meetings. In these meetings, information about ROM and about the study was repeated and the interpretation of the feedback, discussion of the feedback, and use of the case consultation form were re-practiced. Also, questions were answered, and individual cases were discussed, as was the resolution of obstacles. The purpose of these meetings was to train new clinicians and to boost the clinicians' motivation for using ROM and for the study, particularly resolving potential obstacles that might lead to the clinicians' discontinuing their use of ROM. In addition, the primary researcher sent a monthly report to the manager and all of the clinicians in each of the departments in which she indicated response percentages and offered advice about how to improve the use of ROM. As a tangible reward for the team, each department that had a ROM percentage above 80 at the start of treatment received a cake. Also, during the entire duration of the study, a research assistant spent one day a week checking whether the protocol procedures were being followed; whether all of the eligible children, adolescents, and parents were being approached; which patients had agreed to participate; and whether participants' case consultation forms had been returned. At the end of each day, the research assistant sent each secretarial staff feedback about registration errors that had been made. She also sent each clinician who was treating participating children, adolescents, or parents a summary report of the assessments that should be completed and the case consultation forms that should be returned. The results of having checked for registration errors and for missing assessments and case consultation forms were regularly discussed with the managers, so that they could encourage their team members to carry out ROM and to remind them to return the case consultation forms. Finally, a helpdesk was available to address technical problems and to answer questions that the secretaries or the clinicians might have. The helpdesk comprised three staff members with advanced knowledge about the ROM system and ROM procedures. The helpdesk could be reached by either telephone (direct answer) or e-mail (answer within a couple of hours). The number of phone calls and e-mail messages that the helpdesk received varied both within and between days. In general, however, both the secretaries and the clinicians used the helpdesk on a regular basis.

Sample Size and Randomization

Based on the literature, our study was powered to detect an effect size of 0.30 (Cohen's *d*, effect size in the small range). Sample size calculations (with $\alpha=0.05$ and $\text{power}=0.80$) indicated that 144 participants would be needed in each of the conditions in order to detect differences among the three feedback conditions. Based on previous experiences, we expected that about 50% of the participants would cease completing the questionnaires following the baseline assessment. Thus, we aimed to recruit twice as many children, adolescents, and parents as the power analysis indicated (288 per condition, total $N=864$) in order to have the required number of participants in each condition for whom the clinicians would have received feedback during the course of the treatment (i.e., at least 50% of the measurements).

Randomization was conducted at the level of participants. Participants were nested within clinicians, and clinicians were nested within departments. Approximately 170 different clinicians worked at the youth departments during the course of the study and each clinician treated between 0 and 5 participants within each condition. Because participants were assigned to departments (entire teams) instead of clinicians, we were unable to account for the clustering of clients within clinicians. We excluded the possibility of a confounding effect of departments within the randomization. Hence, participants were randomly allocated to one of the three conditions (allocation ratio [1:1:1]), and they were stratified by location (four departments) and participants' age (4–11 year olds and 12–17 year olds). Hereby, a block randomization procedure was used. An independent researcher created the randomization blocks using a computer-generated procedure. The research assistant filled out the randomization blocks during the inclusion period and determined the condition to which each participant was assigned. She also matched the participants to the corresponding condition in the ROM system, so that participants received the web-based questionnaires for that specific feedback condition. At the end of the inclusion period, 343 participants had been randomized (Department 1: 31 participants 4–11 years old and 71 participants 12–17 years old; Department 2: 25 participants 4–11 years old and 72 participants 12–17 years old; Department 3: 15 participants 4–11 years old and 55 participants 12–17 years old; and Department 4: 39 participants 4–11 years old and 33 participants 12–17 years old).

Participants

Figure 1 shows the flow of participants through the study. A total of 1606 children, adolescents, and parents were eligible to participate. A total of 343 children and adolescents were randomized; 55 children and adolescents completed the questionnaires themselves; for 118 children and adolescents, their parents completed the questionnaires; and for 170 children and adolescents, both the child or the adolescent and his or her parents completed the questionnaires. There was a significant difference between eligible and participating children and adolescents in their country of birth ($\chi^2 [1, N=1598]=5.59, p=0.018$). More nonparticipating children and adolescents ($n=57, 4.5\%$) than participating children and adolescents ($n=6, 1.7\%$) were born outside the Netherlands.

The youth sample consisted of 225 children and adolescents with a mean age of 15.08 years ($sd=1.55$), and 139 (61.8%) of them were female. The majority of the children and adolescents ($n=218, 96.9\%$) were born in the Netherlands and lived in

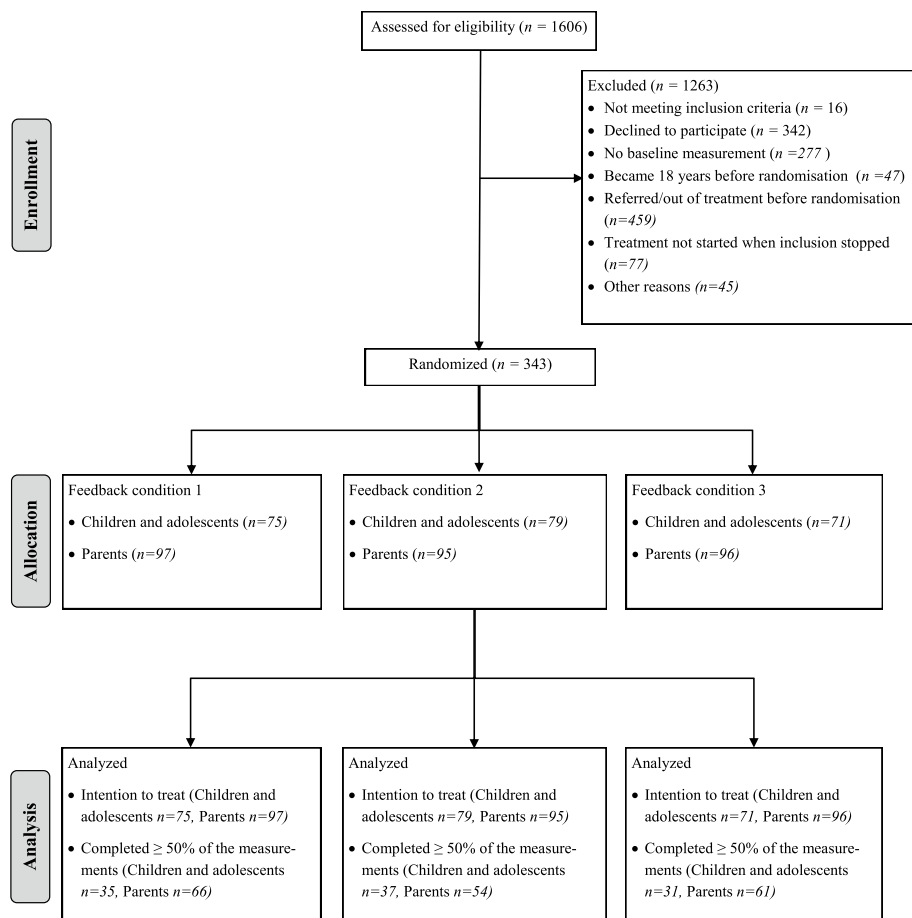


Fig. 1 Participation flow through the study

a two-parent household ($n = 146$, 64.9%). The main diagnosis of the children and adolescents was a neurodevelopmental disorder ($n = 81$, 36.0%), in particular autism spectrum disorder ($n = 38$) or attention-deficit/hyperactivity disorder ($n = 40$); a depressive disorder ($n = 48$, 21.3%); a trauma- and stressor-related disorder ($n = 30$, 13.3%); or an anxiety disorder ($n = 23$, 10.2%). The mean treatment duration for the children and adolescents was 12.25 months ($sd = 5.40$), during which there was an average of 25.5 treatment sessions ($sd = 29.6$).

The parental sample ($n = 288$) comprised 234 mothers (81.3%) and 54 fathers (18.8%). The parents' child or adolescent was on average 12.50 years old ($sd = 3.43$), and 47.2% ($n = 136$) of the children and adolescents were female. The majority of the children and adolescents ($n = 282$, 97.9%) were born in the Netherlands and lived in a two-parent household ($n = 182$, 63.2%). The main diagnosis of the children and adolescents was a neurodevelopmental disorder ($n = 152$, 52.8%), in particular autism spectrum disorder ($n = 72$) or attention-deficit/hyperactivity disorder ($n = 74$); a depressive disorder ($n = 46$, 16.0%); a trauma- or stress-related disorder ($n = 28$, 9.7%); or an anxiety disorder ($n = 28$, 9.7%). The

mean treatment duration for the children and adolescents was 13.44 months ($sd=6.42$), with an average of 20.8 treatment sessions ($sd=23.7$).

Feedback Conditions

In all three of the feedback conditions, the feedback was given to the clinicians, and consisted of a written and graphic summary of the results from the assessment, which were compared to norms and cut-off scores. The feedback was given at the start of treatment, one and one-half months after treatment had started, and subsequently every three months during treatment. The feedback was accessible directly after the questionnaires had been completed in the ROM system and transmitted to the electronic health record one day after the questionnaires had been completed. Thus, the feedback was integrated into the patient's record in which the clinician also annotated the therapy; therefore, it was integrated into the everyday workflow and was readily accessible. However, within the electronic health record system, it was not possible to generate an alert to the clinician whenever new feedback was available.

In the first (control) condition, the clinicians received basic feedback regarding the child's or the adolescent's symptoms and quality of life (from the SDQ and the KID-SCREEN). The feedback included tables and graphs with information about the participant's current total and subscale scores, changes in the scores across time, and a list of items on which the maximum score had been achieved (i.e., critical items that required special attention). By means of this feedback, clinicians were able to evaluate how the child, adolescent, or parent was responding to the treatment.

In the second condition, the feedback that was used in the first condition was extended with CST, which were based on the Treatment Support Measure (TSM; Warren and Lambert 2013). The TSM identifies possible obstacles to a favorable outcome that the child, adolescent, or parent might be facing. The parents of children and adolescents between 4 and 17 years old completed the TSM-Parent Form (TSM-P), and the children and adolescents between 12 and 17 years old completed the TSM-Youth Form (TSM-Y). The TSM-P and TSM-Y each consists of 40 items, which are scored on a five-point Likert scale (ranging from *strongly disagree* to *strongly agree*) and which can be divided into subscales. The TSM-P subscales are *Parenting Self-Efficacy*, *Parent Social Support*, *Parenting Skills*, *Parent Distress*, and *Parent Therapeutic Alliance*. The TSM-Y subscales are *Self-Efficacy*, *Social Support*, *Motivation for Treatment*, and *Therapeutic Alliance*. The TSM was recently validated for use in the Dutch population (van Sonsbeek et al. 2017). For both the TSM-P and the TSM-Y, the internal consistency reliability and convergent validity were good, but divergent validity was less convincing, and the results for criterion validity were inconclusive. Based on the results from the TSM, the clinician received supplementary tables and graphs. These tables and graphs provided information about the current scores and changes in the scores across time. When a subscale score exceeded the cut-off score, practical suggestions were given to the clinician about how the treatment might be improved. Examples of these practical suggestions are: Regularly highlight areas in which the youth has shown improvement and use these successes as leverage to extend self-efficacy to other areas, provide encouragement and support by relapses, role play social situations to facilitate the acquisition of social skills, and discuss clinician and clinical style match. By discussing this feedback, the clinician and the child, adolescent or parent could determine how to improve treatment or get back on track for a good outcome.

In the third condition, the clinicians received all of the information and feedback that was provided in the second condition, but additionally they had to discuss the feedback with a colleague while using a standardized format for case consultation. During the training sessions, all of the clinicians were trained in the use of feedback from ROM and the use of the standardized format for case consultation. The manager of the department was present at the training, expressed support for the study, and emphasized that for children and adolescents in the third condition it was obligatory that the feedback be discussed with a colleague. When a clinician received feedback from a child or adolescent in the third condition, he or she was reminded at the beginning of the feedback to discuss the feedback with a colleague. Then, the clinician chose a colleague to meet with, for example the responsible psychiatrist or the colleague who provided the training in parenting skills. The intention was for the clinician to choose a colleague who had personal relevance for him or her and whose advice he or she would trust. The format for the case consultation was designed so that it could be completed within fifteen minutes and clinicians were accounted for the time they spend with discussing the format by the management. After they had been completed, the case consultation forms had to be returned to the researcher. Thus, in the third feedback condition, the clinician was obliged to pay attention to the feedback and received additional advice from a colleague about ways in which the treatment might be improved. Because a peer who had personal relevance for the clinician gave the additional advice, we expected that clinicians would be more likely to accept the feedback, would become more involved and would feel more responsible to act upon the feedback.

Additionally, we determined whether each child or adolescent was NOT between the first and second assessment. We designated children and adolescents as NOT if their improvement on the total difficulties score on the SDQ from the start of treatment to the first assessment during treatment was less than 8.5% (a detailed description of how NOT is calculated is given at the end of the Analyses section). Whether each child or adolescent was NOT was determined after the inclusion period had ended and was not provided to the clinician.

It proved to be technically impossible to check whether the clinicians had opened the feedback and we were unable to check whether the clinicians had discussed the feedback with the children, adolescents, and parents during the treatment sessions. Consequently, whether the clinicians had discussed the feedback with their clients was confirmed by asking the children, adolescents, and parents about this within the standard set of questionnaires given at each assessment and at the end of treatment (“Did the clinician discuss the feedback from the questionnaires that you filled out [last time]?”). Furthermore, in the third feedback condition, we confirmed whether the clinicians had discussed the feedback with a colleague by checking the case consultation forms that had been returned.

Outcome Measures

Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ; Goodman 1997; Van Widenfelt et al. 2003) measures symptom severity. The children and adolescents between 12 and 17 years old completed the SDQ-S, and the parents of children between 4 and 17 years old completed the SDQ-P. The SDQ consists of 25 items and an additional assessment of impact. All of the items are scored on a three-point scale (ranging from *not true* to *completely true*), which can be divided into the subscales *Emotional Symptoms*, *Conduct Problems*,

Hyperactivity-Inattention, Peer Relationship Problems, and Prosocial Behavior. A total difficulties score can also be calculated by summing the scores on the *Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention, and Peer Relationship Problems* subscales (range of scores: 0–40). The Dutch version of the SDQ-S and the SDQ-P have been found to have acceptable internal consistency, test–retest stability and parent–youth agreement, and good concurrent validity (Muris et al. 2003; Van Widenfelt et al. 2003). However, the reliability of two of the subscales of the SDQ-P (*Conduct Problems* and *Peer Relationship Problems*) has been found to be insufficient (Stone et al. 2010).

KIDSCREEN

The KIDSCREEN-52 child-adolescent version and KIDSCREEN-27 parent version were used to measure the child's or adolescent's quality of life (Ravens-Sieberger et al. 2006). The children and adolescents between 12 and 17 years old completed the KIDSCREEN-52, and the parents of children between 4 and 17 years old completed the KIDSCREEN-27. The items on both the KIDSCREEN-52 and the KIDSCREEN-27 are scored on a 5-point scale ranging from *not at all/never* to *totally/always*. The KIDSCREEN-52 consists of 52 items that can be divided into the subscales *Physical Well-Being, Psychological Well-Being, Moods and Emotions, Self-Perception, Autonomy, Parent Relation and Home Life, Financial Resources, Social Support and Peers, School Environment, and Social Acceptance* (vs. *Bullying*). The KIDSCREEN-52 has acceptable to satisfactory reliability and validity (Ravens-Sieberger et al. 2006, 2008). The KIDSCREEN-27 consists of 27 items that can be divided into the subscales *Physical Well-Being, Psychological Well-Being, Autonomy and Parent Relation, Social Support and Peers, and School Environment*. The KIDSCREEN-27 has satisfactory item internal consistency, item discriminant validity, agreement between youth and proxy reports, and good reliability (Ravens-Siebere).

End of Treatment Variables

The child and parent versions of the Youth Thermometer were used to measure participants' satisfaction with the treatment that they had received (Bransen et al. 2005; Kok and Van Wijngaarden 2003). The Youth Thermometer-Child Version consists of 28 items. The Youth Thermometer-Parent Versions asks about the child's treatment and the training in parenting skills (if applicable). It consists of 31 and 32 items, respectively. The items are answered either *yes/no*, ask for a rating, or have an open-ended response. The items can be divided in the subscales *Appraisal of Information, Appraisal of Participation, Appraisal of the Clinician* (the child's clinician and the parents' clinician), *Appraisal of the Treatment Result* and *Background Information*. The Youth Thermometer-Child Version has acceptable to satisfactory reliability, and the Youth Thermometer-Parent Versions have good reliability (Kok and Van Wijngaarden 2003). The internal consistency of the parent version that asks about the treatment of the child is as yet unclear, but the internal consistency of the parent version that asks about the training in parenting skills is good (Kok and Van Wijngaarden 2003).

The length of treatment was measured as days between admission to and discharge from the mental health care institution and was reported in months. The number of treatment sessions was counted as the total number of contacts between the clinician and the child, adolescent, or parent. The rate of dropout was calculated as the percentage of children and adolescents who terminated the treatment (unilateral decision to end the treatment)

compared to the percentage of children and adolescents who completed the treatment (bilateral decision to end the treatment).

Analyses

To determine whether randomization had resulted in equal distributions across the three feedback conditions in participants' demographic characteristics and their baseline SDQ Total and KIDSCREEN Total scores, we performed one-way ANOVAs and chi-square tests using SPSS (IBM Corp, 2011).

The feedback effects were analyzed both in accordance with the intention-to-treat principle (ITT) and for the completers only (CO). For the intention-to-treat analyses, all of the children and adolescents who had been randomized to one of the three conditions were included (youth sample: $n=225$; parent sample: $n=288$). The children, adolescents, and parents were designated as completers if they had filled out the complete set of questionnaires for at least 50% of the assessments for which they had been invited (youth sample: $n=103$; parent sample: $n=181$). There were no significant differences between the completers and non-completers in their baseline demographic characteristics. We were faced with both participants ceasing to fill out the questionnaires during treatment and incomplete registration of end-of-treatment variables. Therefore, the total number of observations varied from 71 to 225 for the youth sample (see Table 2) and from 79 to 288 for the parent sample (see Table 3).

The primary outcome measure was Symptom Severity as measured with the SDQ. Secondary outcome measures were Quality of Life (which was measured with the KIDSCREEN), and the end-of-treatment variables Satisfaction with Treatment (which was measured with the Youth Thermometer), Length of Treatment, Number of Treatment Sessions, and Rate of Dropout. To test the hypothesis that symptom severity decreased more and quality of life increased more for children and adolescents in the second and third feedback conditions than for children and adolescents in the first condition, we calculated difference scores between the baseline and end-of-treatment assessments. Using the statistical package Mplus 7 (Muthén, and Muthén 1998–2015), we tested whether the three conditions showed significant differences in mean difference scores. The latter was tested by comparing an unconstrained model (in which mean differences are freely estimated in each condition) with a constrained model (in which mean differences are constrained to be equal across conditions) using chi-square difference tests. Missing data were taken into account by using full information maximum likelihood (FIML). Differences among the conditions in the end of treatment variables were also tested with chi-square difference tests in Mplus. Again, we compared unconstrained models (in which means are freely estimated) with constrained models (in which means are constrained to be equal). Differences between participants who dropped out and those who completed treatment were tested using Fisher's exact test in SPSS.

Additionally, we tested whether being NOT moderated the association between the feedback intervention and changes in symptom severity. We planned to determine whether each child or adolescent was NOT by calculating the reliable change index (RCI) for the total difficulties score on the SDQ between the first measurement (baseline assessment) and second measurement (the first assessment during treatment). Children and adolescents were designated NOT if their RCI was smaller than 1.96 (i.e., not showing statistically significant improvement; Jacobson and Truax 1991). However, using this definition we found that 81.2% (from the youth sample) or 82.8% (from the parent sample) of the children and

adolescents turned out to be NOT. Thus, we concluded that this RCI, which is in general used as a measure of outcome at the end of treatment, was too strict to be used as a definition for being NOT during treatment. In addition, by using this definition we found a significant difference in the baseline scores of the OT and NOT groups. This suggested that the NOT group included more children and adolescents with a baseline score near the clinical cut-off score. As a result of having low baseline scores, these children and adolescents were less likely to show statistically significant improvement and thus were more likely to be designated NOT. We decided, therefore, to take the baseline severity score into account in our definition of NOT.

Baseline severity can be taken into account by using percentage improvement (Hiller, Schindler, and Lambert 2012). Percentage Improvement (PI) is calculated as:

$$\frac{[(\text{SDQ} - \text{score at baseline} - \text{SDQ} - \text{score at first assessment during treatment}) / (\text{SDQ} - \text{score at baseline})] * 100.}$$

The amount of change depends on the degree of pretreatment severity. So low baseline severity (i.e., low SDQ scores at baseline) requires relatively small decreases to achieve substantial improvement, whereas high baseline severity (i.e., high SDQ scores at baseline) requires relatively large decreases to attain substantial improvement. The percentage improvement from baseline to the end of treatment that is often used is 50% (Dimidjian et al. 2006 ; Strauman et al. 2006). However, 50% improvement would be too strict for defining NOT during treatment. In an effort to identify a criterion that would be a reasonable estimate for improvement from baseline to the first assessment during treatment, we calculated a mean length of treatment for our sample. The mean duration of treatment for the children and adolescents was 12.25 months (*sd* = 5.40 months; range = 3–30 months). We defined being NOT at the first assessment during treatment, which was on average after 2.07 months. Because the first assessment was administered on average at 17% of the treatment duration, we anticipated finding an improvement of at least 8.5% ($0.17 \times 50\%$) at that point in treatment, with the assumption that improvement was linear across time (Barkham et al. 2006). Thus, we designated children and adolescents as NOT if their improvement on the total difficulties score on the SDQ from the start of treatment to the first assessment during treatment was less than 8.5%. The moderating effect of NOT on the relationship between the feedback condition and changes in severity symptoms was evaluated with regression analyses in Mplus. Feedback condition was represented by two dummy variables; NOT was also represented as a dummy variable; and the two interaction terms were the product terms of the two feedback condition dummy variables and the NOT dummy variable. The predictor variables were the two dummy variables for the feedback condition, the dummy variable for NOT, and the two interaction terms. The dependent variable was change in symptom severity (SDQ total and subscale scores) from baseline to the end of treatment.

Results

Baseline Comparisons

Table 1 summarizes participants' demographic characteristics and their baseline SDQ and KIDSCREEN total scores. In both the youth sample and the parent sample, no significant differences were observed across the three feedback conditions in participants' age, sex,

Table 1 Participants' Baseline Demographics for the Feedback Conditions and Tests for Equality of Means Across Conditions

Baseline Demographics	Condition 1	Condition 2	Condition 3	Comparisons
Youth sample, <i>n</i>	75	79	71	
Age, <i>M (sd)</i>	15.09 (1.59)	15.06 (1.49)	15.08 (1.60)	$F(2, 222) = 0.01, p = .992$
Male, <i>n (%)</i>	26 (34.7)	32 (40.5)	28 (39.4)	$\chi^2(2, N = 225) = 0.62, p = .733$
Country of birth				
Netherlands, <i>n (%)</i>	72 (96.0)	79 (100.0)	67 (95.8)	$\chi^2(4, N = 225) = 6.58, p = .160$
Other country, <i>n (%)</i>	3 (4.0)	0 (0.0)	4 (4.2)	
Household composition				
Two-parent household, <i>n (%)</i>	47 (62.7)	55 (69.6)	44 (62.0)	$\chi^2(10, N = 225) = 8.44, p = .586$
Other, <i>n (%)</i>	28 (37.3)	42 (30.4)	27 (38.0)	
Main diagnosis				
Neurodevelopmental, <i>n (%)</i>	30 (40.0)	30 (38.0)	21 (29.6)	$\chi^2(20, N = 225) = 18.89, p = .529$
Depressive, <i>n (%)</i>	16 (21.3)	13 (16.5)	19 (26.7)	
Trauma- and stressor-related, <i>n (%)</i>	6 (8.0)	14 (17.8)	10 (14.1)	
Other, <i>n (%)</i>	23 (30.7)	22 (27.7)	21 (29.6)	
SDQ-S total, <i>M (sd)</i>	14.99 (5.80)	16.57 (6.14)	16.15 (4.83)	$F(2, 222) = 1.61, p = .202$
KIDSCREEN-52 total, <i>M (sd)</i>	184.21 (33.58)	186.41 (35.23)	177.56 (28.86)	$F(2, 222) = 1.45, p = .236$

Table 1 (continued)

Baseline Demographics	Condition 1	Condition 2	Condition 3	Comparisons
Parent sample, <i>n</i>	97	95	96	$\chi^2 (2, N=288) = 1.09, p = .581$
Mother, <i>n</i> (%)	82 (84.5)	75 (78.9)	77 (80.2)	
Age, <i>M</i> (<i>sd</i>)	12.43 (3.39)	12.60 (3.36)	12.46 (3.56)	$F(2, 285) = 0.07, p = .937$
Male, <i>n</i> (%)	49 (50.5)	47 (49.5)	56 (58.3)	$\chi^2 (2, N=288) = 1.80, p = .406$
Country of birth				
Netherlands, <i>n</i> (%)	93 (95.9)	95 (100.0)	94 (97.9)	$\chi^2 (4, N=288) = 5.99, p = .200$
Other country, <i>n</i> (%)	4 (4.1)	0 (0.0)	2 (2.1)	
Household composition				
Two-parent household	59 (60.8)	63 (66.3)	60 (62.5)	$\chi^2 (6, N=288) = 3.46, p = .750$
Other	38 (39.2)	32 (33.7)	36 (37.5)	
Main diagnosis of child/adolescent				
Neurodevelopmental	55 (56.7)	47 (49.5)	50 (52.1)	$\chi^2 (18, N=288) = 20.90, p = .291$
Depressive	13 (13.4)	13 (13.7)	20 (20.8)	
Trauma- and stressor-related	5 (5.2)	15 (15.8)	8 (8.3)	
Anxiety	13 (13.4)	6 (6.3)	9 (9.4)	
Other	11 (11.3)	14 (14.7)	9 (9.4)	
SDQ-P total, <i>M</i> (<i>sd</i>)	16.76 (7.09)	16.32 (6.22)	17.09 (6.25)	$F(2, 285) = 0.34, p = .712$
KIDSCREEN-27 total, <i>M</i> (<i>sd</i>)	93.14 (15.57)	93.96 (16.32)	92.48 (15.71)	$F(2, 285) = 0.21, p = .812$

country of birth, household composition, primary diagnosis, baseline SDQ total score, or baseline KIDSCREEN total score. Additionally, in the parent sample the number of mothers who were the main responding parent did not differ across the three feedback conditions.

Feedback Effects

The results for the youth sample are shown in Table 2. For the primary outcome measure, symptom severity, there were no significant differences among the three feedback conditions in mean difference scores for both the total and subscales scores. The chi-square difference tests showed no significant change between the unconstrained model (in which mean differences were freely estimated for each condition) and the constrained model (in which mean differences were constrained to be equal across conditions; see final column in Table 2). The results for the secondary outcomes were almost the same.

Table 2 Youth Sample: Differences in Symptom Severity, Quality of life, Satisfaction with Treatment, Length of Treatment, Number of Sessions, and Dropout for the Feedback Conditions

	Condition 1		Condition 2		Condition 3		Comparisons	
	M _{diff}	SD _{diff}	M _{diff}	SD _{diff}	M _{diff}	SD _{diff}	$\sqrt{\chi^2(2)}$	p
<i>Outcome measures</i>								
N at start of treatment	75		79		71			
SDQ total	− 4.07	5.81	− 4.70	4.19	− 5.20	6.11	0.66	0.721
Emotional symptoms	− 0.19	2.53	− 1.77	2.26	− 2.37	2.80	0.88	0.642
Conduct problems	− 0.61	1.35	− 0.93	1.46	− 0.59	1.75	0.33	0.515
Hyperactivity-inattention	− 1.12	2.10	− 1.35	1.78	− 1.47	2.24	0.52	0.770
Peer relationship problems	− 0.62	1.49	− 0.63	1.42	− 0.68	1.52	0.04	0.982
Prosocial behavior	0.35	1.58	0.92	1.57	0.61	1.24	0.32	0.191
Impact	− 2.19	2.94	− 1.22	3.01	− 1.22	3.64	0.07	0.216
KIDSCREEN total	15.49	29.68	15.22	28.83	24.66	28.17	0.29	0.318
Physical well-being	1.67	4.70	0.85	3.07	2.00	3.20	2.43	0.297
Psychological well-being	3.26	4.61	3.22	5.39	4.78	5.43	2.27	0.321
Moods and emotions	3.93	7.89	3.83	6.50	5.51	6.78	1.14	0.566
Self-perception	1.51	3.86	0.65	4.16	2.68	3.60	4.52	0.105
Autonomy	1.69	4.21	1.31	3.79	2.11	3.24	0.89	0.642
Parent relation/home life	0.60	4.67	2.30	4.38	1.34	4.77	3.07	0.216
Financial resources	− 0.11	3.73	0.89	2.74	1.23	2.90	3.37	0.185
Social support and peers	1.99	4.80	1.13	4.65	3.06	4.79	2.76	0.251
School environment	1.48	6.12	2.57	6.63	2.72	7.67	0.84	0.657
Social acceptance	1.48	2.55	0.43	2.13	0.18	2.76	0.52	0.023
<i>End of treatment variables</i>								
	M	sd	M	sd	M	sd		
N for satisfaction with treatment	22		26		23			
N for next two variables	66		71		63			
Satisfaction with treatment	7.82	1.15	8.12	1.25	6.83	2.26	5.46.	0.65
Length of treatment	11.76	5.35	12.49	4.72	12.49	6.06	0.84	0.656
Number of sessions	27.36	37.78	24.65	22.18	24.44	26.68	0.30	0.860
Dropout (number)	1		6		6		Fisher's exact test $p = .106$	

That is, there were no significant differences across the three feedback conditions in changes in quality of life except for *Social Acceptance*. The improvement in *Social Acceptance* from baseline to the end of treatment was significantly higher in Condition 1 than in Condition 2 or Condition 3. There also were no significant differences between the three feedback conditions in the end-of-treatment variables: satisfaction with treatment, length of treatment, number of sessions, or rate of dropout.

The results from the parent sample are shown in Table 3. There were no significant differences among the three feedback conditions in changes in symptom severity or quality of life. There also were no significant differences among the feedback conditions in the end-of-treatment variables: satisfaction with treatment (neither in the treatment of the child nor the training in parenting skills), length of treatment, number of sessions, or rate of dropout.

The results for the ITT group and the CO group were similar for both the youth sample and the parent sample. Therefore, only the results of the ITT group are reported here.

Table 3 Parent Sample: Differences in in Symptom Severity, Quality of life, Satisfaction with Treatment, Length of Treatment, Number of Sessions, and Dropout Between the Feedback Conditions

	Condition 1		Condition 2		Condition 3		Comparisons	
	M _{diff}	SD _{diff}	M _{diff}	SD _{diff}	M _{diff}	SD _{diff}	χ^2 (2)	p
<i>Outcome measures</i>								
N at start of treatment	97	95	96					
SDQ total	− 4.21	4.90	− 4.94	5.44	− 5.19	6.28	1.04	0.594
Emotional symptoms	− 1.70	1.71	− 1.81	2.37	− 1.93	2.55	0.37	0.832
Conduct problems	− 0.91	1.59	− 1.10	1.78	− 0.84	1.66	0.69	0.708
Hyperactivity-inattention	− 0.88	2.17	− 1.21	2.15	− 1.44	2.06	2.01	0.367
Peer relationship problems	− 0.82	1.92	− 0.72	1.53	− 1.01	− 3.82	0.74	0.690
Prosocial behavior	0.43	1.72	0.84	1.47	0.69	1.94	1.92	0.384
Impact	− 1.66	2.47	− 1.90	2.62	− 2.61	3.02	4.00	0.136
KIDSCREEN total	8.81	13.93	9.79	13.51	9.04	13.96	0.17	0.919
Physical well-being	1.67	4.29	1.19	3.73	1.43	3.84	0.42	0.810
Psychological well-being	3.30	3.99	3.48	4.78	3.69	5.15	0.24	0.889
Autonomy	1.46	4.87	1.59	3.05	0.85	3.25	1.60	0.450
Social support and peers	1.33	3.74	1.17	3.23	1.63	3.74	0.55	0.761
School environment	1.43	3.61	2.28	4.14	1.39	4.33	1.96	0.376
<i>End of treatment variables</i>								
	M	sd	M	sd	M	sd		
N for treatment child	40		36		41			
N for parenting skills training	32		23		24			
N for next two variables	84		86		85			
<i>Satisfaction with treatment</i>								
Treatment child	7.08	1.66	7.72	1.17	7.24	1.56	4.53	0.104
Parenting skills training	7.13	1.62	7.74	1.33	7.50	1.12	2.37	0.305
Length of treatment	13.01	7.21	13.51	5.17	13.78	6.66	0.52	0.770
Number of sessions	21.41	31.09	19.86	17.01	21.16	20.78	0.28	0.869
Dropout (number)	2		5		6		Fisher's exact test p = .338	

Moderating Effect of NOT

To determine whether a child or adolescent was NOT, we used an improvement of less than 8.5% on the SDQ total difficulties score between the first assessment (the baseline assessment) and the second assessment (the first assessment during treatment). This resulted in 53.8% of the children and adolescents being NOT in the youth sample and 50.2% of the children and adolescents being NOT in the parent sample. We found no moderating effect of being NOT on the association between the feedback intervention and symptom severity (SDQ total and subscale scores). This finding was similar for the youth sample (SDQ total: for the first interaction term $B = -0.80$, $SE = 2.65$, $p = 0.711$, and for the second interaction term $B = -1.68$, $SE = 2.82$, $p = 0.551$) and the parent sample (SDQ total: for the first interaction term $B = -0.43$, $SE = 2.02$, $p = 0.832$ and for the second interaction term $B = -1.54$, $SE = 2.00$, $p = 0.441$). Again, the findings for the youth sample and the parent sample were similar for the ITT group and the CO group; therefore, results from the latter are not reported here.

Post-Hoc Analyses

We measured implementation by asking children, adolescents, and parents whether the clinicians discussed the feedback within the standard set of questionnaires at each assessment and at the end of treatment. However, only 14.2% of the children and adolescents ($n = 32$; Condition 1 = 16.0%, $n = 12$; Condition 2 = 12.7%, $n = 10$; Condition 3 = 14.1%, $n = 10$), and 22.6% of the parents ($n = 65$; Condition 1 = 24.7%, $n = 24$; Condition 2 = 21.1%, $n = 20$; Condition 3 = 21.9%, $n = 21$) answered the question about whether the clinician discussed the feedback during treatment at the end of their treatment. Only 25.0% of these children and adolescents and 15.4% of these parents reported that the therapist discussed the feedback each time after the questionnaires had been completed. Additionally, 31.3% of these children and adolescents and 32.3% of these parents reported that the feedback was discussed some of the times after the questionnaires had been completed. Thus, it appears that about half of the time the clinicians did not discuss the feedback during the treatment sessions. In the third feedback condition, we additionally measured fidelity to this condition by checking whether the case consultation forms had been returned. For 29.6% ($n = 21$) of the children and adolescents in the youth sample and 36.0% ($n = 35$) of the children and adolescents in the parent sample, the case consultation form was returned at least once during treatment. This raised the question as to whether the results would differ if they were analyzed separately for the children and adolescents with whom the case consultation was used and the feedback was likely discussed during treatment. Because whether there was a case consultation with a colleague defines the difference between the second and third feedback conditions, most of the children and adolescents who were included in the analysis for the third group actually belonged to the second group. We, therefore, included in the second feedback group the children and adolescents for whom a case consultation had not been used. The final comparison included children and adolescents in feedback Condition 1 versus children and adolescents in feedback Condition 2 including the children and adolescents in feedback Condition 3 for whom a case consultation had not been used, as opposed to children and adolescents from feedback Condition 3 for whom a case consultation had been used.

For both the youth sample and the parent sample, we found no significant differences among the three feedback conditions in symptom severity (SDQ total), quality of life (KIDSCREEN total), or the end-of-treatment variables satisfaction with treatment (both treatment of the child and training in parenting skills), length of treatment, number of sessions, and rate of dropout.

Discussion

The present study reports the results of a three-arm, parallel-group, randomized controlled trial, in which a literature-based, multi-faceted implementation strategy was used, that investigated several potentially effective components of feedback from ROM in youth mental health care in the Netherlands. Contrary to expectations, the results showed no significant differences among the three forms of feedback on symptom severity, quality of life, satisfaction with treatment, length of treatment, number of sessions, or rate of dropout. Furthermore, there was no moderating effect of being NOT on the association between the feedback intervention and symptom severity. Research on the use of feedback in youth mental health care is scarce and, as far as we are aware, all available studies have investigated the effects of a specific kind of feedback compared to no feedback or a low intensity of feedback. We are unaware of any other studies that have investigated different kinds of feedback; thus, our results appear to be unique.

The first and most obvious explanation for not finding any differences among the various forms of feedback is that our statistical power was limited. Despite the fact that we extensively informed children, adolescents, parents, and clinicians about the study and we extended the inclusion period, we were unable to recruit the number of participants for which we had aimed. It turned out that a large group of eligible children and adolescents was excluded for participation due to the complex informed-consent procedure. Because parents have legal responsibility for their children, the informed consent form had to be signed by both parents for children between 4 and 12 years old and by both the adolescent and his or her parents for adolescents between 12 and 15 years old. Especially when parents were divorced and when adolescents were in puberty, it was difficult for the informed consent procedure to be completed. Nevertheless, we did not identify any trends in the outcome measures that approached significance.

A second explanation is that the lack of difference among the three feedback conditions was due to poor implementation. Our literature-based, multi-faceted, clinician oriented implementation strategy (Grol et al. 2010) proved insufficient for creating enough trust and involvement in order to effectively implement the use of feedback from ROM. The implementation strategy included training, a ROM implementation package, update meetings, feedback about and rewards for good response rates, reminders to correct errors and to take additional action, and the provision of a helpdesk. In actual practice, the clinicians failed to successfully utilize the feedback and the case consultation. For the feedback to be effective, the questionnaires had to be completed, the clinician had to review the feedback, and the clinician had to discuss the feedback during treatment (Scott and Lewis 2015). We found, however, that only 46% of the children and adolescents and 63% of the parents completed at least 50% of the assessments for which they had been invited. This means that the feedback, on average, was available only half of the time when it was supposed to be. Because it was technically impossible to confirm whether the clinicians had opened the feedback and because we were unable to confirm whether the clinicians

had discussed the feedback, the percentage of the feedback that the clinician viewed and used during treatment was probably even lower. Moreover, only 56% of the children and adolescents and 48% of the parents reported that the clinician discussed the feedback at least once during treatment. This might have created a vicious cycle. That is, only a portion of the children, adolescents, and parents completed the questionnaires; the clinicians either did not discuss why the questionnaires had not been completed or did not discuss the feedback; the children, adolescents, and parents might then have thought that the questionnaires were not important, and so they stopped filling them out. We aimed to evaluate the effects of feedback when both the questionnaires had been filled out and the results had been discussed at least once during a case consultation. The numbers, however, were too small for meaningful conclusions to be reached. An additional complicating factor was that multiple clinicians had been involved in the treatment of a particular child or adolescent. There might have been confusion about which clinician was supposed to discuss the feedback with which participant (child or adolescent and parents), with the consequence that no discussion of the feedback took place. Previous research has consistently confirmed that mental health clinicians only infrequently engage in ROM practices in both adult and child mental health care (e.g., Lyon et al. 2016; Mellor-Clark et al. 2016; Waldron et al. 2018). Specifically, Bickman et al. (2016) and Brattland et al. (2018) showed a direct effect of successful implementation of ROM on treatment outcome. In another study, Bickman et al. (2011) found a dose–response effect, with stronger effects when clinicians reviewed a greater number of feedback reports. It might be that clinicians who fail to use ROM provide treatments that are insufficiently attuned to their clients' problems and preferences, perhaps do not modify treatments when clients are not progressing, and thus have clients with poorer treatment outcomes.

A third explanation for the lack of differences among the three forms of feedback might be that feedback simply does not have an effect on treatment outcome for children, adolescents, and their parents. The small number of previous studies on feedback in youth mental health care have indeed shown a limited effect or no effect at all of feedback on symptom severity, length of treatment, number of sessions, rate of drop-out, or client satisfaction. However, it was not possible for us to evaluate this explanation in the present study, because all of the participating youth departments had already started using ROM, including providing feedback to the clinicians.

Some limitations of the study should be acknowledged. First, caution is needed when attempting to generalize the present findings to other mental health care institutions. Although the participating children, adolescents, and parents were diverse in socio-demographic characteristics, and the children and adolescents had a wide variety of mental health problems, the study was conducted in only one large mental health care institution. In addition, there was a significant difference between the eligible and participating children and adolescents in their country of birth. Second, it is possible that differences in the care that the clinicians provided to the children, adolescents, and parents other than the feedback might have emerged (creating a performance bias), because the clinicians could not be blinded. Because children, adolescents, and parents were assigned to departments (i.e., entire teams) instead of to individual clinicians, we were also unable to account for clustering of clients within clinicians. Third, the children, adolescents, and parents were asked to fill out web-based questionnaires only at the start of treatment, one and a half months after treatment had started, subsequently every three months during treatment, and at the end of treatment. Because clinicians, managers, and directors did not support the idea of session-by-session assessments, we chose initially to implement ROM at a relatively low frequency, to collect more information about the potentially effective components of

ROM, and then to administer the measures more frequently. However, the relatively low frequency at which the outcome measures were administered might have prevented administration of them from becoming routine and therefore a new habit of the clinicians (Nilsen, et al. 2012). Fourth, the extent to which the clinicians actually viewed and discussed the feedback with the children, adolescents, and parents is unclear. It proved technically impossible to confirm whether the clinicians had opened the feedback, and we were unable to confirm whether the clinicians had discussed the feedback with the children, adolescents, and parents during the treatment sessions. Only within the context of administering the standard set of questionnaires were we able to ask the children, adolescents, and parents about whether the feedback had been discussed in their treatment sessions, and only a small percentage of the children, adolescents, and parents responded to this question. Fifth, in investigating whether being NOT moderated the association between feedback condition and changes in symptom severity, we assumed that improvement within treatment would be linear across time. However, in the literature, studies have shown that improvement is both linear and non-linear (e.g. log linear). Although unlikely, it is possible that our results would have differed if we had assumed non-linear improvement. Finally, because of ethical and financial reasons, we were unable to assess potential long-range changes after the treatment had ended.

In conclusion, the fact that we did not find any differences among the various forms of feedback could be attributable to limited power (due to low recruitment), implementation difficulties (such as infrequent ROM, unknown levels of viewing and sharing of feedback reports, or clinicians' lack of adherence to the feedback conditions), or an inadequate theoretical framework (a lack of effect of different forms of feedback on treatment outcome for children, adolescents, and their parents). Because of our power and implementation difficulties, at this time no firm conclusions about the effectiveness of different components of feedback in youth mental health care can be reached, so that the underlying theoretical framework should be retained.

Although this study did have certain limitations, it nevertheless contributes to the limited knowledge base about feedback in youth mental health care generally and outside the United States and for young children specifically. The results, therefore, are relevant for both clinicians and researchers. On the one hand, the results might suggest that more extended forms of feedback and the addition of case consultation do not have incremental effects on treatment outcomes for children, adolescents, and their parents. The results might also suggest that being NOT does not moderate the association between the feedback intervention and changes in symptom severity. On the other hand, we did encounter power and implementation difficulties, and this underscores the complexity of conducting research on and implementation of ROM in youth mental health care in the real world. Complicating factors included the informed consent procedure for children younger than 16 years old, the involvement of both the child or the adolescent and one or both parents in the treatment, the involvement of multiple clinicians in the treatment of a particular child or adolescent, and unexpected obstacles, such as changes in the clinical and the managerial staff. Clearly, additional research is needed in which additional implementation efforts are undertaken.

A recent systematic review (Forman-Hoffman et al. 2017) showed that there is sparse evidence to support the use of diverse implementation strategies. Both researchers and clinicians still do not have adequate knowledge about the best way to successfully implement evidence-based practices in clinical settings for children and adolescents. The evidence is inconsistent for strategies that use educational meetings or educational materials. There is, however, modest evidence to support the use of financial incentives,

such as pay for performance. Implementation strategies appear to be most successful when they include outreach visits and reminders, provide practitioners with newly collected clinical information, and focus on organizational changes. Besides having specific implementation strategies, having more time for implementation seems important. Research has consistently shown that acquiring the ability to effectively use new clinical procedures takes three to five years (e.g., Fixsen et al. 2005). A recent study in adult mental health care also found that the impact of feedback increased across time (Brattland et al. 2018). Little effect of feedback was found in the first and second years of its use, but by the fourth year clients were two and a half times more likely to improve when their clinicians used feedback. Thus, to successfully implement ROM and the use of feedback from ROM in youth mental health care, a number of components seem to be essential, including greater simplicity of procedures, greater stability in the staff, financial incentives for clinicians, outreach visits, continual stimulation for clinicians, a general focus on organizational efficiency, and more time for implementation.

Additional research about the effects of feedback from ROM is obviously needed in youth mental health care, outside the United States, with young children and their parents, with different forms of feedback, and with case consultation. Future research should also identify the mechanisms through which feedback might work, for example by including measures of clinician characteristics and the degree of match between a clinician and a particular child or adolescent. Additionally, it would be important for future studies to have a higher ROM frequency, be longitudinal, and include multiple follow-up assessments after treatment has ended. Also, future research should be designed in such a way that it reduces the bias due to participant and clinician awareness that a specific ROM component is being used in the intervention (Kendrick et al. 2016). Because it is impossible to totally blind participants and clinicians, designs that vary the amount and timing of the feedback would be useful (Chang et al. 2012). Future studies should also find ways to reduce the relatively high rates of attrition. This might be accomplished through independent assessments of research outcomes by staff who are not involved in treating the participants and who have the ability to collect data on participants who do not return. Lastly, future research should explicitly focus on implementation of the feedback intervention, by paying extensive and continuing attention to participants, clinicians, and organizational processes, in order to optimize the chances of identifying potential feedback effects. An effectiveness-implementation hybrid design (Curran et al. 2012), in which components of clinical effectiveness and implementation research designs are blended, might help by paving the way for more rapid translation of research findings into clinical practice, more effective implementation strategies, and eventually more useful information and clear-cut results.

Acknowledgements We would like to thank all of the children, adolescents, parents, and clinicians who participated in the study. We would also like to thank Femke Scheffers for her assistance with the data collection.

Author Contributions MS, GH, JWV and BT were responsible for the study conception and design. MS was responsible for the data collection. Analysis were performed by MS, MK, and AV. MS wrote the first draft of the manuscript, and all authors edited and revised portions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded in part by a grant from ZonMw, the Dutch organization for Health research and Development (Grant Number 159010001/Project Number 80-82470-98-006-08).

Availability of Data and Material The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics Approval The study was approved by the Ethics Committee of Radboud University's Faculty of Social Sciences (ECG2012-1304-031) and performed in accordance with the national medical ethical guidelines and ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Consent to Participate Informed consent was obtained from all participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barkham, M., Connell, J., Stiles, W. B., Miles, J. N., Margison, F., Evans, C., et al. (2006). Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology, 74*(1), 160.
- Bergman, H., Kornør, H., Nikolakopoulou, A., Hanssen-Bauer, K., Soares-Weiser, K., Tollefsen, T. K., et al. (2018). Client feedback in psychological therapy for children and adolescents with mental health problems. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD011729.pub2>.
- Bickman, L., Douglas, S. R., De Andrade, A. R. V., Tomlinson, M., Gleacher, A., Olin, S., et al. (2016). Implementing a measurement feedback system: A tale of two sites. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 410–425.
- Bickman, L., Kelley, S. D., & Athay, M. (2012). The technology of measurement feedback systems. *Couple and Family Psychology: Research and Practice, 1*(4), 274–284.
- Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*(12), 1423–1429.
- Bransen, M., Kok, I., & Van Wijngaarden, B. (2005). *De aanpassing van de GGZ jeugdthermometer versie 2003 tot de GGZ jeugdthermometer versie 2005*. Utrecht: Trimbos-instituut.
- Brattland, H., Koksvik, J. M., Burkeland, O., Gråwe, R. W., Klöckner, C., Linaker, O. M., et al. (2018). The effects of routine outcome monitoring (ROM) on therapy outcomes in the course of an implementation process: A randomized clinical trial. *Journal of Counseling Psychology, 65*(5), 641.
- Carlier, I. V., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of Evaluation in Clinical Practice, 18*(1), 104–110.
- Chang, T. E., Jing, Y., Yeung, A. S., Brennenman, S. K., Kalsekar, I., Hebden, T., et al. (2012). Effect of communicating depression severity on physician prescribing patterns: findings from the Clinical Outcomes in measurement-based Treatment (COMET) trial. *General Hospital Psychiatry, 34*, 105–112. <https://doi.org/10.1016/j.genhosppsych.2011.12.003>.
- Claiborn, C. D., & Goodyear, R. K. (2005). Feedback in psychotherapy. *Journal of Clinical Psychology, 61*(2), 209–217.
- Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care, 50*(3), 217.


- De Jong, K., Timman, R., Hakkaart-Van Roijen, L., Vermeulen, P., Kooiman, K., Passchier, J., et al. (2014). The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychotherapy Research*, 24(6), 629–639.
- de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research*, 22(4), 464–474.
- Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmalings, K. B., Kohlenberg, R. J., Addis, M. E., et al. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, 74(4), 658.
- Fixsen, D., Naoom, S., Blase, K., Friedman, R., & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tamps, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.
- Forman-Hoffman, V. L., Middleton, J. C., McKeeman, J. L., Stambaugh, L. F., Christian, R. B., Gaynes, B. N., et al. (2017). Quality improvement, implementation, and dissemination strategies to improve mental health care for children and adolescents: a systematic review. *Implementation Science*, 12(1), 93.
- Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(3), 325–343.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586.
- Grol, R., Wensing, M., & Implementatie: Effectieve verbetering van de patiëntenzorg., (2010). *Maars-sen*. Nederland: Elsevier Gezondheidszorg.
- Hamilton, J. D., & Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(10), 1114–1119.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61(2), 155–163.
- Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, 17(1), 25–32.
- Harmon, C., Hawkins, E. J., Lambert, M. J., Slade, K., & Whipple, J. S. (2005). Improving outcomes for poorly responding clients: The use of clinical support tools and feedback to clients. *Journal of Clinical Psychology*, 61(2), 175–185.
- Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research*, 22(1), 1–11.
- IBM Corp. Released (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.
- Ivers, N. M., Tu, K., Francis, J., Barnsley, J., Shah, B., Upshur, R., et al. (2010). Feedback GAP: study protocol for a cluster-randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care. *Implementation Science*, 5(1), 98.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Jensen-Doss, A., Haimes, E. M. B., Smith, A. M., Lyon, A. R., Lewis, C. C., Stanick, C. F., et al. (2018). Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 45(1), 48–61.
- Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., et al. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD011119.pub2/abstract>.
- Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *The British Journal of Psychiatry*, 195(1), 15–22.
- Kok, I., & Van Wijngaarden, B. (2003). Cliëntwaardering in de GGZ. Vragenlijsten en handleiding: GGZ Thermometer voor de volwassenenzorg (versie 2003) en GGZ Jeugdthermometer (versie 2003).

- Krägeloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. *Psychiatric Services*, 66(3), 224–241.
- Lambert, M. J., Bailey, R., Kimball, K., Shimokawa, K., Harmon, S. C., & Slade, K. (2007). *Clinical support tool manual-brief version-40*. Salt Lake City: OQ Measures.
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55(4), 520.
- Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., et al. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA psychiatry*, 76(3), 324–335.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Upper Saddle River: Prentice-Hall.
- Lyon, A. R., Ludwig, K., Wasse, J. K., Bergstrom, A., Hendrix, E., & McCauley, E. (2016). Determinants and functions of standardized assessment use among school mental health clinicians: A mixed methods evaluation. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(1), 122–134.
- Mellor-Clark, J., Cross, S., Macdonald, J., & Skjulsvik, T. (2016). Leading horses to water: Lessons from a decade of helping psychological therapy services use routine outcome measurement to improve practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(3), 279–285.
- Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology*, 61(2), 199–208.
- Muris, P., Meesters, C., & van den Berg, F. (2003). The strengths and difficulties questionnaire (SDQ). *European child & Adolescent Psychiatry*, 12(1), 1–8.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide (Version 7)* (p. 2004). Los Angeles, CA: Muthén & Muthén.
- Nilsen, P., Roback, K., Broström, A., & Ellström, P. E. (2012). Creatures of habit: accounting for the role of habit in implementation research on clinical behaviour change. *Implementation Science*, 7(1), 53.
- Østergård, O. K., Randa, H., & Hougaard, E. (2018). The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy—A systematic review and meta-analysis. *Psychotherapy Research*, 1–18.
- Patterson, P., Matthey, S., & Baker, M. (2006). Using Mental Health Outcome Measures in Everyday Clinical Practice. *Australasian Psychiatry*, 14(2), 133–136.
- Ravens-Sieberer, U. (2006). *The Kidscreen questionnaires: quality of life questionnaires for children and adolescents; handbook*. Lengerich: Pabst Science Publ.
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., et al. (2008). The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Value in health*, 11(4), 645–658.
- Reese, R. J., Duncan, B. L., Bohanske, R. T., Owen, J. J., & Minami, T. (2014). Benchmarking outcomes in a public behavioral health setting: Feedback as a quality improvement strategy. *Journal of Consulting and Clinical Psychology*, 82(4), 731.
- Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. *Social Psychology and Evaluation*, 104, 138.
- Scott, K., & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice*, 22(1), 49–59.
- Seidman, E., Chorpita, B. F., Reay, W. E., Stelk, W., Garland, A. F., Kutash, K., et al. (2010). A framework for measurement feedback to improve decision-making in mental health. *Administration and Policy in Mental Health and Mental Health Services Research*, 37(1–2), 128–131.
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298.
- Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research*, 22(6), 638–647.
- Slade, K., Lambert, M. J., Harmon, S. C., Smart, D. W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, 15(5), 287–303.
- Sparks, J. A., & Duncan, B. L. (2018). The partners for change outcome management system: A both/and system for collaborative practice. *Family Process*, 57(3), 800–816.

- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical Child and Family Psychology Review*, 13(3), 254–274.
- Strauman, T. J., Vieth, A. Z., Merrill, K. A., Kolden, G. G., Woods, T. E., Klein, M. H., et al. (2006). Self-system therapy as an intervention for self-regulatory dysfunction in depression: A randomized comparison with cognitive therapy. *Journal of Consulting and Clinical Psychology*, 74(2), 367.
- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(4), 547.
- Tam, H. E., & Ronan, K. (2017). The application of a feedback-informed approach in psychological service with youth: Systematic review and meta-analysis. *Clinical Psychology Review*, 55, 41–55.
- van Sonsbeek, M. A., Holtmaat, C. J., Tiemens, B. G., Hutschemaekers, G. J., & de Jong, K. (2017). Psychometric properties of the dutch version of the treatment support measure (TSM) parent and youth form. *Clinical Psychology & Psychotherapy*, 24(5), 1205–1218.
- van Sonsbeek, M. A., Hutschemaekers, G. G., Veerman, J. W., & Tiemens, B. B. (2014). Effective components of feedback from routine outcome monitoring (ROM) in youth mental health care: study protocol of a three-arm parallel-group randomized controlled trial. *BMC psychiatry*, 14(1), 3.
- Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, 12(6), 281–289.
- Waldron, S. M., Loades, M. E., & Rogers, L. (2018). Routine outcome monitoring in CAMHS: how can we enable implementation in practice? *Child and Adolescent Mental Health*, 23(4), 328–333.
- Walfish, S., McAlister, B., & O'donnell, P., & Lambert, M. J., (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, 110(2), 639–644.
- Warnick, E. M., Gonzalez, A., Robin Weersing, V., Scahill, L., & Woolston, J. (2012). Defining dropout from youth psychotherapy: How definitions shape the prevalence and predictors of attrition. *Child and Adolescent Mental Health*, 17(2), 76–85.
- Warren, J. S., & Lambert, M. J. (2013). *Youth clinical support tools manual*. Treatment support measure: Parent and youth forms.
- Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*, 78(2), 144.
- World Health Organization. (2018). Adolescent mental health - Factsheet. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Maartje A. M. S. van Sonsbeek^{1,2}  · **Giel J. M. Hutschemaekers**^{1,3,4} · **Jan W. Veerman**⁴ · **Ad Vermulst**⁵ · **Marloes Kleinjan**^{6,7} · **Bea G. Tiemens**^{1,3,4}

¹ Pro Persona Research, Pro Persona, Postbus 27, 6870 AA Renkum, The Netherlands

² Pro Persona Youth, Pro Persona, Renkum, The Netherlands

³ Indigo Gelderland, Indigo Service Organisatie BV, Utrecht, The Netherlands

⁴ Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

⁵ GGZ (Mental Health Care) Oost Brabant, Boekel, The Netherlands

⁶ Department of Epidemiology, Trimbos Institute, Utrecht, The Netherlands

⁷ Department of Interdisciplinary Social Sciences, Utrecht University, Utrecht, The Netherlands